

Copyright (c) 1993 - 2000 Compugen Ltd.	Gencore version 4.5		
OM protein - protein search, using sw model			
Run on:	March 1, 2001, 15:52:17 ; Search time 170.72 Seconds (without alignments) 17.102 Million cell updates/sec		
Title:	US-09-331-631A-1_COPY_74_116		
Perfect score:	248		
Sequence:	1 NQEDFQTECQQCQRRCRQOE..... RQQYCQRCKEICEEEBYY 43		
Scoring table:	BLOSUM62		
Gapop 10.0 , Gapext 0.5			
Searched:	195891 seqs, 6790655 residues		
Total number of hits satisfying chosen parameters:	195891		
Minimum DB seq length:	0		
Maximum DB seq length:	2000000000		
Post-processing:	Minimum Match 0% Maximum Match 100% Listing first 45 summaries		
Database :	PIR_666;*		
1:	pir1;*		
2:	pir2;*		
3:	pir3;*		
4:	pir4;*		
Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.			
SUMMARIES			
Result No.	Score	Query Match Length DB ID	Description
1	117	47.2 566 2 S22477	vicilin precursor
2	105	42.3 588 1 FWCNAB	alpha-globulin B p
3	104	41.9 509 2 S08059	alpha-globulin typ
4	96	38.7 411 2 T29475	hypothetical prote
5	94	37.9 605 2 S06398	alpha-globulin typ
6	80	32.5 425 2 T1592	hypothetical prote
7	80	32.5 600 2 T18593	hypothetical prote
8	80	32.5 810 2 T4430	protein PV100 lmp
9	79	31.9 572 2 T29890	hypothetical prote
10	78	31.5 47 2 JC5557	arginine/glutamate
11	77.5	31.2 242 2 T2699	hypothetical prote
12	74	29.8 388 2 T31887	hypothetical prote
13	74	29.8 388 2 T31888	hypothetical prote
14	74	438 2 T31889	hypothetical prote
15	73.5	29.6 335 2 T31560	hypothetical prote
16	73.5	29.6 335 2 T31561	hypothetical prote
17	72.5	29.2 330 2 T25169	hypothetical prote
18	72.5	28.8 335 2 T31559	hypothetical prote
19	71.5	28.8 1513 2 T23681	hypothetical prote
20	70	28.2 654 2 T30136	hypothetical prote
21	68	27.6 69 2 T18450	hypothetical prote
22	68	27.4 419 2 T18450	hypothetical prote
23	67	27.0 154 2 T29667	hypothetical prote
24	66.5	26.8 273 2 T20930	hypothetical prote
25	66.5	26.8 314 2 T27686	hypothetical prote
26	66.5	26.8 356 2 S31574	hypothetical prote
27	66.5	26.8 637 2 S35221	globulin Bg1 prec
28	65	26.2 111 2 T20452	hypothetical prote
29			androgen receptor
2			androgen receptor
A34721			hypothetical prote
A3948			antimicrobial pept
T13592			protamine - mouse
A41822			protamine 1 - rat
S03997			brain-specific mem
J01150			protein kinase (EC
T14004			trifA protein - sII
KASHS2			alpha-S2-casein pr
T08875			histidine kinase h
T03454			AIR protein - huma
T23682			hypothetical prote
JE0120			glycoprotein A - m
146207			involuterin - dog
ALIGNMENTS			
RESULT	1		
S22477			vicilin precursor - cacao
C-Species: Theobroma cacao (cacao)			
C-Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 17-Mar-1999			
C-Accesision: S22477; S22478; S18105; S22050			
R-MCHENRY, L.; FRITZ, P.J.			
Plant Mol. Biol. 18, 1173-1176, 1992			
A-Title: Comparison of the structure and nucleotide sequence of vicilin genes of coco			
A-Reference number: S22477; MUTDB:92288309			
A-Accesision: S22477			
A-Cross-references: EMBL:X62626			
A-Molecule type: DNA			
A-Residues: 1-566 <MCH>			
A-Cross-references: EMBL:X62625			
A-Accession: S22478			
A-Molecule type: mRNA			
A-Residue: 1-452 <MC2>			
A-Cross-references: EMBL:X62626			
A-GeneID: 1566			
A-Cross-references: EMBL:X62625			
A-Accession: S22478			
A-Molecule type: mRNA			
A-Residue: 1-452 <MC2>			
A-Cross-references: EMBL:X62626			
A-GeneID: 1566			
F-1-24/Domain: signal sequence #status predicted <SIG>			
F-25-566/Product: vicilin #status predicted <MAT>			
Query Match	47.2%	Score 117; DB 2; Length 566;	
Best Local Similarity 50.0%; Pred. No. 4 2e-05;			
Matches . 20; Conservative 10; Mismatches 10; Indels 0; Gaps 0;			
QY 2 QDPQTSCQQCQRRCRQESGPSPQQDQYCORRCKEICEEE 41			
Db 78 EERBLQRYQQGCRCONQQSGREQQQCQRCWEQQE 117			
RESULT	2		
FWCNAB			
alpha-globulin B precursor (clone C72) - upland cotton			
N-Alternate names: seed storage protein; vicilin precursor			
C-Species: <i>Gossypium hirsutum</i> (upland cotton)			
C-Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 16-Jul-1999			
C-Accesision: A30838; S06911			
R-CHLAN, C.A.; PYLE, J.B.; LEGOCKI, A.B.; DURE III, L.			
Plant Mol. Biol. 7, 475-489, 1986			
A-Title: Developmental biochemistry of cottonseed embryogenesis and germination XVIII			
A-Reference number: A30838			
A-Accesision: A30838			
A-Molecule type: mRNA			
A-Residues: 1-588 <CHL>			
A-Cross-references: GB:M16891; NID:q167374; PID:AAA33071.1; PID:q167375			
A-Experimental source: Var. Coler 201			
R-CHLAN, C.A.; BORROTO, K.; KAMALAY, J.A.; DURE III, L.			

Plant Mol. Biol. 9, 533-546, 1987
A;Title: Developmental biochemistry of cottonseed embryogenesis and germination. XIX. Se
A;Reference number: S06398
A;Accession: S06398
A;Status: not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 1-81 <CH2>
C;Comment: This is a seed storage protein.
C;Superfamily: glycinin
C;Keywords: glycoprotein; seed; storage protein
F;1-25/Domain: signal sequence #status predicted <SIG>
F;26-588/Product: alpha-globulin storage protein #status predicted <MAT>
F;417/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;417/Binding site: carbohydrate (Asn) (covalent) #status predicted <MAT>

Query Match 38.7%; Score 96; DB 2; Length 411;
Best Local Similarity 39.5%; Pred. No. 0.0048;
Matches 17; Conservative 11; Mismatches 11; Indels 4; Gaps 2;
Matches 19; Conservative 9; Mismatches 10; Indels 2; Gaps 1;

Qy 2 QEDPQTECQQ-CORRCQEQESPRQQYCQRKKEICEEEERY 43
Db 138 OOPAQOCOCOACQQQQ---OPQQOCOCOTCQSDDQY 177

RESULT 3
S08059 alpha-globulin type B precursor (tandem 1) - upland cotton (fragment)
N;Alternate names: seed storage protein
C;Species: *Gossypium hirsutum* (upland cotton)
C;Date: 31-Mar-1990 #sequence_revision 31-Mar-1990 #text_change 30-Sep-1993
C;Accession: S08059
R;Chlam, C.A.; Borroto, K.; Kamalay, J.A.; Dure III, L.
Plant Mol. Biol. 9, 533-546, 1987
A;Title: Developmental biochemistry of cottonseed embryogenesis and germination. XIX. Se
A;Reference number: S06398
A;Accession: S08059
A;Status: not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 1-1509 <CH2>
C;Superfamily: glycinin

Query Match 39.5%; Score 105; DB 1; Length 588;
Best Local Similarity 42.3%; Pred. No. 0.0074;
Matches 19; Conservative 11; Mismatches 10; Indels 2; Gaps 1;

Qy 3 EDPQTECQQCORRCQEQESPRQQYCQRKKEICEEEERY 42
Db 81 EDPQRYEECQQECRQEE--RQPOCQQRCLKRFEQQ 118

RESULT 4
T18592 hypothetical protein AC3.3 - *Caenorhabditis elegans*
C;Species: *Caenorhabditis elegans*
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jan-2000
C;Accession: T18592
R;McMurray, A.
submitted to the EMBL Data Library, April 1996
A;Reference number: 218995
A;Accession: T18592
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-425 <WIL>
A;Cross-references: EMBL:271177; PIDN:CAA94867.1; GSPDB:GN00023; CESP:AC3.3
C;Genetics:
A;Gene: CESP:AC3.3
A;Map position: 5
A;Introns: 18/3
C;Superfamily: gliadin

Query Match 32.5%; Score 80.5; DB 2; Length 425;
Best Local Similarity 37.9%; Pred. No. 0.19;
Matches 17; Conservative 9; Mismatches 12; Indels 23; Gaps 2;

Qy 2 QEDPQTECQQCORRC-----QEQESGP-----RQQYCQRKKEICE 38
Db 255 QQPSSAPQCQCOQNTCQQAAPCQQCAPCQQQTSCQTQCQQCTPQO 314
A;Gene: CESP:T01D1.6
A;Map position: 2
A;Introns: 25/3; 304/3
C;Superfamily: gliadin

RESULT 5
S06398 alpha-globulin type A precursor - upland cotton
N;Alternate names: seed storage protein
C;Species: *Gossypium hirsutum* (upland cotton)
C;Date: 31-Mar-1990 #sequence_revision 31-Mar-1990 #text_change 30-Sep-1993
R;Chlam, C.A.; Borroto, K.; Kamalay, J.A.; Dure III, L.
Plant Mol. Biol. 9, 533-546, 1987
A;Title: Developmental biochemistry of cottonseed embryogenesis and germination. XIX. Se
A;Reference number: S06398
A;Accession: S06398
A;Status: not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 1-605 <CH2>
C;Superfamily: glycinin
F;1-24/Domain: signal sequence #status predicted <MAT>
F;25-605/Product: alpha-globulin type A #status predicted

Query Match 37.9%; Score 94; DB 2; Length 605;
Best Local Similarity 35.9%; Pred. No. 0.01;
Matches 14; Conservative 12; Mismatches 13; Indels 0; Gaps 0;

Qy 2 QEDPQTECQQCORRCQEQESPRQQYCQRKKEICEEEERY 40
Db 115 QQPDKOFKECQRCQWQEPRKQOCVKCREQYQED 153

RESULT 6
T18592 hypothetical protein AC3.3 - *Caenorhabditis elegans*
C;Species: *Caenorhabditis elegans*
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C;Accession: T18592
R;McMurray, A.
submitted to the EMBL Data Library, April 1996
A;Reference number: 218995
A;Accession: T18592
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-425 <WIL>
A;Cross-references: EMBL:271177; PIDN:CAA94867.1; GSPDB:GN00023; CESP:AC3.3
C;Genetics:
A;Gene: CESP:AC3.3
A;Map position: 5
A;Introns: 18/3
C;Superfamily: gliadin

Query Match 32.5%; Score 80.5; DB 2; Length 425;
Best Local Similarity 37.9%; Pred. No. 0.19;
Matches 17; Conservative 9; Mismatches 12; Indels 23; Gaps 2;

Qy 2 QEDPQTECQQCORRC-----QEQESGP-----RQQYCQRKKEICE 38
Db 255 QQPSSAPQCQCOQNTCQQAAPCQQCAPCQQQTSCQTQCQQCTPQO 314
A;Gene: CESP:T01D1.6
A;Map position: 2
A;Introns: 25/3; 304/3
C;Superfamily: gliadin

RESULT 7
T18593 hypothetical protein AC3.4 - *Caenorhabditis elegans*
C;Species: *caenorhabditis elegans*
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T18593
R;McMurray, A.
Submitted to the EMBL Data Library, April 1996
A;Reference number: Z18995
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-600 <WIL>
A;Cross-references: EMBL:Z71177; PIDN:CAA94868.1; GSPDB:GN00023; CESP:AC3.4
A;Experimental source: clone AC3
C;Genetics:
A;Gene: CESP:AC3.4
A;Map position: 5
A;Introns: 292/2; 374/3; 572/2

RESULT 8
T44430 protein PV1.00 [imported] - winter squash
C;Species: *Cucurbita maxima* (winter squash)
C;Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jul-2000
C;Accession: T44430
R;Yamada, K.; Shimada, T.; Kondo, M.; Nishimura, M.; Hara-Nishimura, I.
J. Biol. Chem. 274, 2563-2570, 1999
A;Title: Multiple functional proteins are produced by cleaving Asn-Gln bonds of a single
A;Reference number: 222767; MUID:9107919
A;Accession: T44430
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-810 <YAM>
A;Cross-references: EMBL:AB019195; NID:93808061; PIDN:BAA34056.1; PID:93808062

Query Match 32.5%; Score 80.5; DB 2; Length 600;
Best Local Similarity 27.9%; Pred. No. 0.25; Mismatches 9; Indels 17; Gaps 2;
Matches 17; Conservative 9; Mismatches 12; Indels 23; Gaps 2;

QY 2 QEDPPQTQQQCORCR-----QOESGP-----ROQQYCORRCKEICE 38
Db 122 QQPSSAPQCQQCQCONTQQQAPVQCCQCAPQCQQCOTSCOOTQCQQCTPQCO 181

QY 39 E 39
Db 182 Q 182

RESULT 10
JC557 arginine/glutamate-rich 6.5k polypeptide - smooth loofah
C;Species: *Luffa cylindrica* (smooth loofah)
C;Date: 02-Sep-1997 #sequence_revision 05-Sep-1997 #text_change 07-May-1999
C;Accession: JC557
R;Kimura, M.; Park, S.S.; Sakai, R.; Yamasaki, N.; Funatsu, G.
Biosci. Biotechnol. Biochem. 61, 984-988, 1997
A;Title: Primary structure of 6.5-k-arginine/glutamate-rich polypeptide from the seeds
A;Reference number: MUID:9735743
A;Accession: JC557
A;Molecule type: protein
A;Residues: 1-47 <KIM>
A;Experimental source: seed
C;Comment: This protein is a storage protein which provides nitrogen and carbon reser
F;12-33, 16-29/disulfide bonds: #status predicted

RESULT 11
T29699 hypothetical protein F31A3.1 - *Caenorhabditis elegans*
C;Species: *caenorhabditis elegans*
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 17-Mar-2000
C;Accession: T29699
R;Murray, J.; Le, T.T.
Submitted to the EMBL Data Library, May 1996
A;Description: The sequence of *C. elegans* cosmid F31A3.
A;Reference number: Z20667
A;Accession: T29699
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-242 <MUR>
A;Cross-references: EMBL:058742; PIDN:AAB36856.1; GSPDB:GN00028; CESP:F31A3.1
A;Experimental source: strain Bristol N2; clone F31A3
C;Genetics:
A;Gene: CESP:F31A3.1
A;Map position: X
A;Introns: 18/3; 160/3
C;Superfamily: ultra-high-sulfur keratin

RESULT 9
T29880 hypothetical protein F19G12.7 - *Caenorhabditis elegans*
C;Species: *Caenorhabditis elegans*
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C;Accession: T29880
R;Nhan, M.; Le, T.T.
Submitted to the EMBL Data Library, April 1996
A;Description: The sequence of *C. elegans* cosmid F19G12.
A;Reference number: Z20704
A;Accession: T29880
A;Status: preliminary; translated from GB/EMBL/DDBJ

Query Match 31.2%; Score 77.5; DB 2; Length 242;
Best Local Similarity 33.3%; Pred. No. 0.26; Mismatches 17; Indels 11; Gaps 4;
Matches 17; Conservative 12; Mismatches 11; Indels 11; Gaps 4;

